Pragmatics 1:1.71-106
International Pragmatics Association
DOI: 10.1075/prag.1.1.04boi

TRANSCRIPTION DESIGN PRINCIPLES FOR SPOKEN DISCOURSE RESEARCH

John W. Du Bois

1 Introduction

Transcription is theory.[1] This inevitable conclusion has been brought home ever more forcefully in the years since Ochs' seminal article (1979), as our awareness of the tangible shape of discourse has sharpened, and as the push to comprehend language in relation to use has whetted appetites for more and more new discourse material—transcribed with a perspicacity and insight that can nourish new theories with the vital information they need to grow. How we transcribe doesn't just reflect our theories of language, it also shapes them, drawing our eyes to some phenomena while leaving others in shadow. We know this, and yet surprisingly little has been written that might help discourse theorists think about how to get a discourse transcription to do what they want it to do (but see Ochs 1979, Edwards 1989 and forthcoming, Edwards and Lampert forthcoming). Partly this involves assessing what a discourse transcription needs to do, and partly it involves figuring out how to frame a system that can do it.

These are the problems I wish to address here. First, I will consider how people use discourse transcriptions, in order to determine what features need to be built into them. Second, I will offer a framework for designing discourse transcription systems to meet these demands, including demands engendered by modern research tools like the computer, and by the nature of their human users. Because transcription is shaped by theory, and discourse theories vary, it is not possible to detail here every viable transcriptional category. Rather, it is my purpose to present general principles for system design. My remarks may be of interest to three kinds of users: those who wish to design a new transcription system to meet their own unique research needs and theoretical orientation; those who wish to adapt an existing system that almost, but not quite, meets their needs; and those who would simply like to know more about the foundations of

transcription system design, in order to deepen their understanding of the transcribing process and the data records which result from it. Discourse transcriptions are becoming *the* key data for many researchers, not only in linguistics but in neighboring disciplines as well. Given the propulsion that new conceptions of data can bring to ongoing theoretical developments, a close examination of discourse transcription from the theoretical groundwork up has value for all participants in the expanding spectrum of spoken discourse studies.

Though the focus will remain on general principles, these must be exemplified through specific illustrations, for which I will draw mainly on a transcription system which I and my colleagues have developed (Du Bois et al. forthcoming a, b). Many of the points could of course be made as well with other good transcription systems (see Edwards and Lampert, forthcoming and the references therein).

2 What should a discourse transcription be?

The first thing to do in designing a discourse transcription system is to assess one's goals and plans for using the transcriptions. What is discourse transcription? Who will use the transcription? What for? What features will best serve the users' goals? In this section I set out general assumptions, goals, and desiderata for discourse transcription, in order to lay the groundwork for the more specific design principles treated in the following section.

What is discourse transcription? Discourse transcription can be defined as the process of creating a representation in writing of a speech event so as to make it accessible to discourse research (Du Bois et al. forthcoming a). What it means to make the event "accessible to discourse research" will of course depend on what kinds of research questions one seeks to answer. Although speech events are always viewed through the lens of some theory, one can try to ensure that the theory—the framework for explanation and understanding—does justice to the spoken reality, or rather, to selected aspects of it. The process of discourse transcription is never mechanical, but crucially relies on interpretation within a theoretical frame of reference to arrive at functionally significant categories, rather than raw acoustic facts (cf. Ochs 1979, Ladefoged 1990, Du Bois et al. forthcoming a).

The nature of discourse transcription is necessarily shaped by its end. Why transcribe? Transcription documents language use, but language use is attested equally in written discourse, which has the advantage of being easy to obtain without transcribing. What makes speaking worth the extra effort? Spoken

language differs in structure from written language, in ways that remain surprisingly little studied: many aspects of spoken grammar, meaning, and even lexicon remain to be documented. Moreover, it is in spoken discourse that the *process* of the production of language is most accessible to the observer. Hesitations, pauses, glottal constrictions, false starts, and numerous other subtle evidences observable in speech but not in writing provide clues to how participants mobilize resources to plan and produce their utterances, and to how they negotiate with each other the ongoing social interaction. Prosodic features like accent and intonation contour provide important indicators of the flow of new and old information through the discourse (Chafe 1987, etc.). And the moment-by-moment flux of speech displays a rich index of the shifting social interactional meanings that participants generate and attend to, as well as of the larger dimensions of culture embodied in social interactions (Goodwin 1981, etc.). A transcription of spoken discourse can provide a broad array of information about these and other aspects of language, with powerful implications for grammar, semantics, pragmatics, cognition, social interaction, culture, and other domains that meet at the crossroads of discourse.

But discourse transcription cannot be equated with simply writing down speech, because there is not, nor ever can be, a single standard way of putting spoken word to paper. An oral historian, a phonetician, a journalist, and a dialectologist will all produce very different renderings of the same recording, and a discourse researcher's transcription will differ yet from all of these. Indeed, because these other methodologies for writing speech are available, the discourse transcriber can defer to them when necessary, relying on the relevant specialist, equipped with the appropriate analytical framework and notational conventions, to deal with those features which are specific to his or her domain of inquiry. This is not to say that the domain of discourse is insulated from, say, phonetics,[2] since clearly a small detail of pronunciation can ironically reverse a conveyed meaning, or even signal a speaker's alignment with one conversational participant against a third. But at least the predictable regularities of phonetics, phonology, grammar and lexicon can generally be assumed to have been described elsewhere (perhaps even in a descriptive grammar written by the same researcher wearing another hat), so that these facts need not be recapitulated in the discourse transcription. The discourse transcription is designed for answering some but not all questions about speaking, and will necessarily contain both more and less information than another discipline's representation of the same event.

Pursuing the nature of discourse transcription further leads to more specific questions: What kinds of events will be transcribed? Who will use the

transcriptions? How will they use them? What should go into a transcription? I take up each question in turn.

What kinds of speech events will be transcribed? A conversation, classroom lecture, committee meeting, political speech, service encounter, or even just a few words exchanged hastily in the hallway might form the object of scrutiny. Each of these speech event types presents somewhat different demands, but a good transcription system should be able to accommodate, with adaptation if need be, the full range of events that are likely to be of interest. In many respects the most challenging case is the free-wheeling multi-party conversation, and any system that can meet its vast demands will have passed the severest test, and positioned itself well to handle other speech events that may be encountered.

Who will use the transcriptions? Discourse researchers, of course, in all their variety. But these days their interest in discourse is shared by an ever-widening circle. Grammarians and general linguists use transcriptions as sources of linguistic data on a range of topics, and to follow the action in theories grounded in discourse; computational linguists use them to test speech recognition protocols against actual language use; language teachers use them to illustrate realistic uses of spoken language; social scientists use them for understanding the nature of social interaction; curious folks find it intriguing to look closely at how people really talk; and the students of any of these may use transcriptions to learn more about their field of study. And, as we shall see, one of the most important groups of users is the transcribers themselves. A good transcription system should be flexible enough to accommodate the needs of all of these kinds of users.

How will people use the transcriptions? The most fundamental thing they do is to read them, perhaps browsing through a transcription (or a stack of them) to look for a particular phenomenon or pattern, or to formulate a hypothesis. This requires the transcription not only to present the needed information but to present it in a way that is easily assimilated. Second, many users will want to search the transcription using a computer in conjunction with various kinds of data management software, which may include a word processor, a database manager, and a concordance maker, among other things. For this, the transcription should make all the necessary distinctions in ways that ensure that searches will be exhaustive and economical (see §3 below and Edwards 1989, forthcoming). It would be hard to overestimate the impact of the microcomputer on discourse transcription design; many of the possibilities, and many of the constraints, that are spoken about in this article would not exist if the computer had not in recent years made itself such an indispensable tool for many, though certainly not all, types of discourse research. This is not to say that the needs of

the discourse researcher should be bent to the requirements of the machine; as Edwards has persuasively argued (1989), and as I will reaffirm below, an aware and purposeful pursuit of certain basic design principles can insure from the outset that it is the computer that adjusts to the needs of the researcher. Finally, one key function that is often overlooked is embedded in the transcribing process itself. Through the experience of transcribing the transcriber is constantly learning about discourse, not only gaining skill in discriminating the categories implicit in the transcription system but also acquiring a vivid image of the conversational reality that he or she is seeking to represent. To the extent that there is more going on than the transcription system can capture, it is the transcriber, immersed in the recorded speech event and grounded in discourse theory, who is in a position to rectify this, to advance the potential of the transcription system and its theoretical framework. Although transcribing is sometimes thought of as a kind of manual labor, merely a necessary means of producing certain valuable end products, in reality the process itself has tremendous potential for enlightening its practitioners, and for generating the level of keen perception and intimate knowledge that can translate into theoretical insight and new research directions. With this in mind, the transcription system should contribute to making the transcribing process a valuable experience in itself. The system should be convenient and comfortable to use, reasonably easy to learn, and through its implicit categories it should promote insightful perception and classification of discourse phenomena, which in the end may feed back into advances in the system itself.

The fact that a transcription is likely to be exploited in several different ways using diverse tools means that it must be transportable (§3.2), that is, it should be straightforward to move data from one context of use to another. As noted, the computational tools of choice may include word processors, concordance-makers, database managers, and other programs, used on systems ranging from IBM PC-compatibles to Macintoshes to mainframes. And the output from these tools—search results, key-word-in-context concordances, the basic transcriptions themselves, and so on—need to be presented in formats as varied as computer text files, database files, screen displays, printouts, handwritten notes, blackboard inscriptions, and printed articles. These transcriptions may come to be read by a wide range of people interested in discourse, from linguists to sociologists to students of foreign languages, in addition to the discourse specialists who may form their initial audience. What this all boils down to is that discourse data need to be represented in a form that is sufficiently robust to maintain data integrity across contexts of use. Practically speaking this means using the most widely standardized symbols available (§3.2, §3.3).

What should go into a transcription? Allowing for theoretical diversity, we can still observe that many transcription systems take pains to represent (or at least make notations available for) many or all of the following features: the words spoken, written so as to allow each lexical item to be recognized; an identification of the speaker of each turn; the temporal sequencing of utterances, whether these follow each other in succession or are simultaneous (as when speakers overlap); basic units in which the utterances were articulated, such as turns and intonation units; intonation contour, whether functionally or phonetically classified; accent; fluctuations in timing such as tempo, pause and lengthening; nonverbal noises made by speech event participants, such as laughter, throat-clearing, inhalation; special qualities of voice that extend over a stretch of speech; non-utterance events that become relevant to the interaction, such as a participant's serving food, or a sudden thunderclap in the background; metatranscriptional and "evidential" comments on the transcription itself, indicating where the transcriber is uncertain of the words spoken, and so on; and other features as appropriate. (Depending on the language, these may include interlinear glosses and free translations, among other things.) Most of this information is locally sequenced—linked to specific moments in the stream of speech—but some of it pertains rather to the whole of the speech event or transcription, and can be presented separately in a header of "global", non-sequenced information containing, for example, general contextual data on the speech event participants, the situation, the event's recording, and its transcription (Du Bois forthcoming).

It will be useful at this point to look at a brief example of a discourse transcription that takes these factors into consideration. The following example is based on the system presented in *Discourse Transcription* (Du Bois et al. forthcoming a, referred to hereafter as DT). It should be borne in mind that DT is only one of many possible systems that could meet the kinds of needs discussed in this article. In the example below,[3] speaker identification is indicated by a colon following the name or label in capital letters; overlap between speakers is indicated by square brackets, with the left brackets vertically aligned; two separate cases of overlap in close succession are distinguished from each other by single versus double brackets; pauses of various lengths are written as clusters of three dots, with the duration of longer pauses given in parentheses; truncation of words is shown by a hyphen at the end of the word; intonation units are marked by carriage return, that is, each line contains a single intonation unit; and intonation contours are broadly classified as to their "transitional continuity" (Du Bois et al. forthcoming a, b) by the comma, period, or question mark at the end of the line. A summary of these and other symbols used in the DT system is given in

Appendix 1. (For a full discussion of each of these transcriptional categories, see Du Bois et al. forthcoming a, b.)

(1) ((CARSALES))

```
D: we have the used Dodge,
   but no new ones.
G: ... No new ones?
   [Huh],
D: [Dodge] is separate,
   from Chrysler Plymouth.
G: ... You're [kidding me].
D:            [I mean] they're all affiliated,
   but th- th- [[they're not]] on our lot.
G:             [[Right]].
   ...(.8) They're not on your lot.
   [Wow].
D: [Yeah].
```

A "narrower" transcription within the same system would add more detail, using additional conventions. In the following illustration, a very short pause (less than 200 milliseconds) is indicated by a cluster of two dots; primary and secondary accent are marked with caret and grave accent, respectively; and laughter is written with one @ symbol for each "syllable" of laughter.

(2) ((LUNCH))

```
L: .. But `they never `figured ^out what he had?
R: ... He had ^pneumonia.
   ... [The ^second `week] he had `pneu@monia,
M:     [^Eventually].
```

These brief illustrations are enough to provide some common ground for the discussion ahead; additional transcription conventions and examples from this system will be introduced as needed.

3 Transcription design principles

With the goals of discourse transcription in mind, we are now in a position to specify the design principles which they motivate. While transcription design principles are commonly left implicit, it is to the benefit of all concerned with the quality of discourse data and theory that they should be made explicit.

We can group our transcription design principles under five broad maxims:

1. *Category definition*: Define good categories.
2. *Accessibility*: Make the system accessible.
3. *Robustness*: Make representations robust.
4. *Economy*: Make representations economical.
5. *Adaptability*: Make the system adaptable.

I will now take up each of these maxims and the design principles they motivate.

3.1 Category definition

DEFINE GOOD CATEGORIES. The most basic issue in designing a discourse transcription system is not choosing symbols, but defining the analytical categories for which the symbols will stand.

1. *Define transcriptional categories which make the necessary distinctions among discourse phenomena.* Deciding on what discourse phenomena should be represented by distinct transcriptional categories is logically the crucial foundational task. It is also the task which is most sensitive to the theoretical assumptions and research goals of the transcription system designer (see §1, §2). For a discussion of category definitions for one transcription system, see Du Bois et al. (forthcoming a, b).

2. *Define sufficiently explicit categories.* Categories should be differentiated at a sufficiently fine delicacy to ensure that searches are economical, and not undermined by false selection. Categories should also be orthogonal (Edwards 1989). The idea is to avoid false finds, for example, coming up with examples of *only* when attempting to find all instances of *on* (Edwards 1989). This reflects the need for a kind of "splitting": making sure that the notation allows sufficient discrimination and doesn't lump categories that ought to kept distinct. For example, some transcription systems use a single symbol, the colon (:), to indicate both prosodic lengthening and speaker identification; but this is likely to produce searches for one category that turn up unwanted instances of the other. In DT, speaker identification is marked by colon (cf. principle 4) while prosodic lengthening is marked by equal sign, which avoids this kind of problem.

(3) ((AFRICA))
B: Did you ever hear that?
A: No=,

Of course, if two different uses of a character always occur in distinct contexts (i.e. are in complementary distribution), then as long as the contexts can

be made explicit it should be possible to ensure that searches will select only the relevant items. Thus DT exploits the slash character / both in the marking of phonetic transcription (/ /) and rising pitch /, since in the first case the paired slashes are always immediately adjacent to a single parenthesis while in the second case this configuration never occurs. There is a distinction here between an ambiguous symbol and a character that happens to form part of two different symbols.

3. *Define sufficiently general categories.* Categories should encompass a sufficient range to ensure that the sets they define will be meaningful, and not partial or distorted. One can think of this as a "lumping" tendency, in a positive sense. This is done to ensure exhaustive selection of all the instances of the category one is seeking (Edwards 1989).

Obviously there is a tension between the two principles just articulated which needs balancing. Edwards (1989) argues that splitting actually presents a greater danger than lumping, since when extraneous instances turn up in a computer search they can be weeded out by hand, but when relevant instances are just not there the researchers may never know what they are missing.

4. *Contrast data types.* As I have noted elsewhere (Du Bois forthcoming), discourse transcriptions typically include a number of distinct types of data, which are of necessity intermingled in a fairly complex fashion, given the intrinsically multi-layered complexity of speech events themselves. It is very useful to visibly contrast these data types, so that readers of a transcription will know at every moment what kind of information they are taking in. Some of the relevant distinctions are: utterances versus their attribution to a speaker; speaking versus non-utterance actions (such as eating); events versus the temporal unfolding of those events; words spoken by speech event participants versus words interjected as commentary by the transcriber; commentary about a specific moment in the speech event versus commentary about the whole event (e.g. about the event setting); and others. While no practical transcription system is likely to draw a rigid visual distinction between all of these types of information, it is important to mark clearly at least the most crucial distinctions.

Perhaps the most familiar illustration of a highly effective differentiation of data types is to be found in the practice of Western playwrights and their publishers. Their conventions for capturing the unfolding of the spoken drama, evolved over centuries of theatrical tradition, neatly mark differences between spoken words, speaker identification labels (or "speech assignments"), actor directions, stage directions, and other data types in a way that can be readily

absorbed at a glance (cf. Catron 1984:83-89). While script formatting conventions vary somewhat from publisher to publisher, as well as across time and across publishing genres, there is one popular format, commonly used in "collected works" editions because of its economy of space, which shows many affinities to popular discourse transcription formats. Consider the following example:

(4) (Wilde 1982:334)
JACK: Gwendolen!
GWENDOLEN: Yes, Mr. Worthing, what have you got to say to me?
JACK: You know what I have got to say to you.
GWENDOLEN: Yes, but you don't say it.
JACK: Gwendolen, will you marry me? (*Goes on his knees.*)
GWENDOLEN: Of course I will, darling. How long you have been about it!

Here the speaker identification labels are written in small capitals ("JACK"), while the words actually spoken are written in mixed-case, i.e. a mixture of upper- and lower-case letters ("Gwendolen!"). This contrast is reinforced by the placement of the colon: words before the colon generally represent speaker labels, while words after the colon generally represent words actually spoken. These two data types are in turn contrasted with a third, the playwright's interjected descriptions of actions, marked by italics enclosed within parentheses ("(*Goes on his knees.*)"). While it would be attractive to simply incorporate all these familiar (cf. principle 5) and effective devices into one's discourse transcription system, for various reasons notations like small capitals and italics are problematic for discourse transcription systems (cf. principle 2, §3.2, and §3.3). But with only slight adaptation, equivalent contrasts between data types can be marked using more suitable symbolic resources. Consider the following discourse transcription example:

(5) ((AFRICA))
```
ALICE:  and I mean,
        ((SNAPS FINGERS)) you can get up,
        just that quick.
        if [you just] stumble and fall.
BETTY:     [Uhhuh],
```

Here, the distinction between speaker identification labels and actual utterances is marked by using capitals ("ALICE:") versus mixed-case ("and I mean"). This traditional contrast is reinforced by the use of colon, with essentially the same meaning in DT as in the playwright's practice (cf. example 4). Similarly, DT distinguishes interjected transcriber commentary from actual speaker utterances via a version of the playwright's use of parentheses-plus-distinctive-font (e.g. italics): DT makes the parentheses double (so as to unambiguously distinguish interjected commentary from nonverbal vocal noises, marked by single parentheses; cf. example 12), and the distinctive "font" employed is simply upper-case letters. This visually reinforces the contrast between transcriber commentary or action descriptions[4] ("((SNAPS FINGERS))") and speaker utterances, written with mixed-case letters. Such data type discrimination should be pursued wherever feasible, within the scope allowed by the somewhat limited typographical resources available, and recognizing the rigorous demand for ambiguity avoidance that the transcription system designer must face.

3.2 Accessibility

MAKE THE SYSTEM ACCESSIBLE. The task of symbolically representing the chosen transcriptional categories should be guided by the need to make the notational system maximally accessible, in terms of both learning and actual use, to users at all levels, including new learners.

5. *Use familiar notations.* Transcription systems have a lot to gain from drawing on existing traditions for representing speech in writing, whenever viable conventions can be found. Oldest and most familiar are literary traditions like those of playwrights and novelists, who have grappled with the problem of representing speech in writing for many centuries. Most discourse transcription systems have borrowed freely from these systems, whether consciously or not.

For example, when writers of fiction in the Western tradition seek to capture the rhythm of spoken interaction, one of the notational devices they often exploit is a sequence of three dots to mark a pause:

(6) (Lodge 1984:272)
"I mean there's something I might say to you which might imply that I didn't trust you."
"What is it?"
"It's ... hard to say."

If this convention is incorporated into a discourse transcription system, anyone who has read a few novels in the Western literary tradition will effortlessly acquire and apprehend the intended meaning:

(7) ((CARSALES))
```
D: I want to make fifty thou a year,
G: ... @ But what about all those phone numbers.
```

Another feature of conversational interaction that novelists are attuned to is truncation (or aposiopesis), often represented by an em dash:

(8) (Hillerman 1990:37)
"For example, from the Zuni sorcery tradition. Or the Hopi 'two-heart' legends, or—" Dr. Bourebonette stopped, midphrase. She looked embarrassed.

The essence of truncation—stopping in midphrase—is even spelled out in words here. This convention has been around for a long time, as suggested by the following example, which is far from the oldest that could be cited:

(9) (Disraeli 1904:93)
'You speak of—' said Egremont, hesitatingly.

Because of its familiarity, DT incorporates this literary convention for truncation into the discourse transcription system (approximating the em dash with a sequence of two hyphens):

(10) ((J&J))
```
J: ... And he --
   and he kicks my feet apart,
```

The playwright's need to display spoken interaction in a way that can be instantly absorbed by actors and readers (example 4) has led to devices and conventions tested by many years of use, whose clarity and simplicity make them well worth borrowing:

(11a) (Williams 1971:162)
AMANDA: Come back here, Tom Wingfield! I'm not through talking to you!
TOM: Oh, go—
LAURA [*desperately*]: —Tom!
AMANDA: You're going to listen, and no more insolence from you! I'm at the end of my patience!
[*He comes back toward her.*]

(11b) (Williams 1971:210-11)
JIM: Hello there, Laura.
LAURA [*faintly*]: Hello.
[*She clears her throat.*]
JIM: How are you feeling now? Better?
LAURA: Yes. Yes, thank you.

The playwright's conventions may well be the richest source of time-tested and familiar notational devices, many of which go back to Shakespeare's time and beyond (e.g. Shakespeare 1623). They contribute powerfully not only to the discrimination of data types (principle 4) but also to the overall display of information on the page (principle 18).

Cartoonists go farther than novelists and perhaps even playwrights in their attempts to capture the expressive character of speech. To be sure, they often draw on special representational resources that neither novelists nor discourse transcribers can count on (see §3.3 below), such as variable size, shape, alignment and even color of characters, as well as balloons and other speech attribution devices (Du Bois 1986). But some of their devices can be carried over, like that of writing nonverbal vocal sounds—such as a sigh or a burp—within parentheses, when these are included in a speech balloon:

(12)
(SIGH) IF GOD HAD WANTED US TO GROW UP, HE WOULDN'T HAVE GIVEN US TEDDY BEARS. (*Momma*, by Mell Lazarus, Los Angeles Times 5/27/80)

Single parentheses here serve to reinforce the distinction between verbal and nonverbal, helping the reader to immediately recognize that (SIGH) indicates not an utterance of the word "sigh" but an actual instance of the action of sighing. In discourse transcription it is equally useful to draw this distinction, for which DT and several other systems borrow the cartoonists' convention.

(13) ((BRIDGE))

```
V: (H) and the only thing going through my mind was,
   (SWALLOW) (TSK) (H) Well,
   I mean,
   I can really kick and fight,
```

Even existing discourse transcription systems, despite their relatively brief history, can sometimes offer notational conventions boasting a degree of familiarity, at least among those already initiated into spoken discourse research. For example, the use of square brackets (or similar symbols) for overlapping speech is familiar to many through exposure to transcriptions in the Conversation Analysis tradition (Atkinson and Heritage 1984); and many transcription systems, including DT, have adopted some sort of bracket-like symbol for this function.

(14) ((DINNER))

```
A: I'm [thinking in] terms of the [[genetic]] factor,
B:     [She was] --
                                  [[She was]] --
C: No.
B: [She was] --
S: [She was] radioactive.
B: She was nuts.
```

(While DT adopts the bracket-like symbol, it takes a slightly different approach to its placement, as discussed below under principle 18.)

Transcription systems benefit greatly from drawing on readers' years of exposure to notations like these for pause, truncation, nonverbal vocal noises, data type discriminations, and so on, because they invoke conventions that cost users nothing to learn, since they already know them. Naturally the means employed by writers to represent speech are not without limitations, which include a degree of ambiguity and inconsistency in usage (Du Bois et al. forthcoming a), as well as a certain lack of expressive power. For example, in a novel three dots do not indicate how long the pause lasts, and sometimes they are used with other meanings such as ellipsis. Similarly, em dash indicates not only truncations but hesitations and parenthetical comments as well. But this merely underscores the task of the transcription system, which is to make explicit, consistent, and unambiguous the meaning attached to every notation—in accordance with transcription design principles like those presented in this article—and to increase expressive power by offering additional conventions.

6. *Use motivated notations.* In addition to the authority accrued historically via tradition (as in the previous section), a notation may be motivated in other ways, such as iconicity and internal consistency. For example, writing a stroke sloping upward to the right (/) for rising pitch and a stroke sloping downward (\) for falling pitch is partly iconic (though also partly grounded in conventions of left-to-right reading order).[5] And aligning speech overlaps vertically, so that words which two speakers begin uttering at the same time are written beginning at the same column on the page, counts as a kind of diagrammatic iconicity. In a similar vein, once it is pointed out that the @ sign for laughter in DT somewhat resembles the widespread "smiley face" symbol (☺), this strikes some people as iconic.[6]

(15) ((DINNER))

```
A: and if you just sort of rinse the spoon off afterwards,
   you don't really [have to wash dishes],
B:                  [@ @@@]
```

The equal sign (=) for lengthening in DT (cf. example 3 above) comes to look like stretch marks on an elastic object being stretched out—at least once you know what it is supposed to mean. (This use of the equal sign derives from a long-standing literary convention, the similarly iconic use of dashes to indicate elongated sounds: "No—o-o-o, I—" (Williams 1971:222).)

Iconicity can be diagrammatic as well, as when the iteration of a symbol four times (e.g. four @ signs) indicates four iterations of an event (e.g. four pulses of laughter). Diagrammatic iconicity facilitates the meaningful use of space (cf. principle 18), making it easy to refer to specific moments within a string of iterated or extended events. For example, when an extended sequence of laughter is partially overlapped, the left overlap bracket can be located precisely at the point in the middle of the sequence of iterated events where the overlap actually begins (and similarly for the right bracket that marks overlap ending).

(16) ((AFRICA))

```
B:  Well,
    I heard of a elephant,
    that sat down on a [VW,
    one time].
A:                     [@] @@@@@@[[@]]
B:                                [[There was a]] girl- --
    Did you ever hear that.
A:  @ No,
```

7. *Use easily learned notations.* In a sense this is simply the cognitive counterpart of motivation. To the extent that new notations must be introduced which will require learning, it is motivated notations (iconic, mnemonic, or internally consistent) that should be learned more quickly. And a transcription system that is easy to learn is more likely to be used, and used without error, than one that is not. For example, the letter H for audible breathing is, if not directly iconic, mnemonic for anyone who reads a language in which orthographic H spells a similar sound.

Ease of learning is also favored if all members of a notational paradigm have the same symbolic syntax, as it were, making the system internally motivated (Du Bois 1985). For example, in DT the brackets that indicate special qualities of voice extending over a stretch of speech always carry their distinctive labeling in the same place. Thus in the symbols for laughing while speaking (**<@ @>**), the tags are attached to both left and right angle brackets on the inside edge. Because the same pattern is used for the DT notations for whispering (**<WH WH>**), quotation quality (**<Q Q>**), codeswitching (**<L2 L2>**), indexed overlap brackets (**[3 3]**), and so on, once the pattern is learned for one symbol it will carry over to the rest of the symbols in the system.

(17) ((AFRICA))

```
A:  they'd even got out of the car,
    He really knew better.
    than to <@ get out of the car @>.
B:  Well,
```

Similarly, pattern congruity argues for marking long pauses with a sequence of dots like the medium and short pauses, despite the fact that simple distinctiveness could be achieved just by the use of numbers in parentheses.

(18) ((RANCH))

```
R: ... This .. is the type of person,
   ...(.9) that ...(.7) is like ...(1.0) a hermit.
```

8. *Segregate unfamiliar notations.* Traditional literary conventions, however useful, cannot by themselves provide enough familiar resources to construct an entire discourse transcription system, so a certain number of unfamiliar and even arbitrary conventions will be called for. But their impact can be lessened if they are visually segregated from the (familiar) representations of words spoken. This can be done in several ways. Restricting categories like laughter and primary accent to non-alphabetic symbols like @ and ^ (as in DT) allows the nonspecialist struggling to follow a narrow transcription to get by by ignoring all the nonalphabetic symbols and reading just the words uttered, which come through loud and clear in their traditional mixed-case alphabetic notation. Over time, the beginner can learn to attend also to the other streams of information carried by the nonalphabetic symbols embedded among familiar words. Even veteran users of transcriptions find it useful to be able to attend to these different data types in visually segregated channels (principle 4), as in the following narrow transcription sample:

(19) ((LUNCH))

```
R: ...(1.0) When they 'quit going to ^Littleton, \
   .. every 'week to see his ^gra@ndmo@the@r @@@, /
   (H)
L: .. Oh 'that's [when he ^outgrew it]? /
R:               [@@@@]
```

Considering the amount of prosodic information conveyed here, this display is significantly more readable than it would be if all information were presented within just the mixed-case "channel", rather than having the prosody (mostly) segregated in a separate non-alphabetic "channel".

9. *Use notations which maximize data access.* The need for diverse tools and formats to serve the various uses of discourse transcriptions has been treated at some length in §2. Clearly some notations are flexible enough to allow access through all of these means, while others are too dependent on one particular tool (for example a word processor with special font capabilities). Notations which are readily transportable (see §3.3 below) not only make the transition easier, they help ensure data integrity.

10. *Maintain consistent appearance across modes of access.* This in some sense follows from the last principle; but when it comes time to publish transcriptions in a book or article it is tempting to throw off the irksome limitations of the ASCII character set in order to draw on the wealth of special symbols, fonts, and formatting possibilities that typesetting can offer. But this has

certain disadvantages, perhaps the most important being that time spent reading the book version and gaining familiarity with its notational conventions will not necessarily carry over to the transcriptions as displayed in other formats. In contrast, maintaining the same conventions and visual display whether the medium is a blackboard, notebook, computer screen, printout, or book will mean that experience does carry over from one medium to the next. Another consideration is that maintaining the same conventions lessens the likelihood that errors will be introduced in the process of converting from one format to another. The argument for using the same conventions across formats applies not only to different media but also to different tools within the same medium, such as word processing versus database software, where the latter may be unable to handle the sophisticated symbols and formats produced by the former. Of course, there is no harm in occasionally drawing on the additional resources of a more flexible medium like print for functions like highlighting speech event features of special interest; for example, DT suggests using boldface, an otherwise nonviable notational resource (cf. principle 11), for highlighting in examples cited in publications (cf. principle 21). It's just that both the integrity of the data and the investment of the learner will be better served if such embellishments are restricted to redundantly enhanced presentation of a stable set of conventions.

3.3 Robustness

MAKE REPRESENTATIONS ROBUST. Once the goals of maximizing data access and transportability across platforms are adopted, this immediately invokes a need for data representations which are robust enough to survive the moves with all their integrity intact. The maxim of robustness, and hence the principles outlined in this section, can thus be considered as corollaries to principles 9 and 10.

11. *Use widely available characters.* Characters which are widely available do not have to be translated, which removes the danger of mistranslation. The most widely available standard is the lower ASCII (7-bit) character set. This puts a severe cramp on the number of symbols that can be drawn on, demanding a certain amount of ingenuity from the transcription designer. On the bright side, it instills discipline and discourages a frivolous proliferation of notations, which would make things more complicated for the learner.

Nothing said here should be taken to imply that people who are already comfortably using non-ASCII characters for writing a given language should stop using them in their discourse transcriptions; it is only that a transcription system designed for general (including international) use should not impose *additional*

demands of this type on its users. For example, the specialist in Swedish or Finnish discourse finds nothing intimidating about venturing beyond the limited horizons of lower ASCII, since such symbols are already in daily use for writing the standard literary language itself. But the transcription system should not presume to add more demands in this direction, especially considering that speakers of different languages around the world are not likely to have the same set of extra-ASCII symbols at their disposal.[7]

As a corollary of this principle, one should avoid using notational resources which are not *standardly represented* across platforms, such as boldface, italics, underlining, special fonts (especially proportional fonts), margin shifts, and so on, as the *sole* marker of crucial contrasts between transcriptional categories.[8] Although many word processors allow boldface, for example, they don't represent it in the same way, so that distinctions can be obliterated or garbled when moving data across platforms. Also, many databases and concordance programs may not be able to effectively handle attributes like italics and boldface.

12. *Avoid invisible contrasts.* If two categories are distinguished by characters that cannot be visually discriminated in all formats, this makes it hard for users to monitor them—for example, in proofreading a paper printout of a computer file—and hence may make it difficult to maintain data integrity across all modes of access. For example, the contrast between a tab and a space (or between a tab and five spaces) should not be used even if the computer screen can display it, because it won't show on paper. In the same way, contrasts based on nonprinting "control" characters create problems for proofing of printouts, and hence for data integrity.

13. *Avoid fragile contrasts.* By "fragile" contrasts I mean those which are easily disrupted or modified in the normal course of working with a transcription. For example, some transcription systems indicate speech overlap by vertically aligning the two speakers' overlapping words, and then aligning a pair of brackets on a third line between them to mark the beginning and ending of the overlap. But if a user changes the margins, tab settings, or fonts, the vertical alignment of these three lines is likely to shift. At best this kind of accidental change will produce a garbled display; but in the worst case, it will result in a configuration that looks like a plausible speech overlap, allowing the spurious alignment to go undetected and uncorrected. Also, moving the data from a word processing file into, say, a database file is likely again to cause vertical alignment information to be modified or lost. In general, contrasts which depend entirely on vertical alignment between separate lines, or on the insertion of sequences of multiple spaces, are prone to unnoticed modifications and should be avoided.[9]

But there is no harm in *supplementing* a secure robust notation with fragile resources like vertical alignment, as long as these are used redundantly. For example, in DT brackets are first placed *within* the line of overlapped speech, where they are not subject to accidental shifts; but once the overlap is securely marked in this way, temporal sequencing and simultaneity can be displayed more iconically by adding vertical alignment. The embedded brackets ensure that any reformatting will not result in *undetectable* changes.

(20) ((DINNER))

```
A: ... What were you doing before.
C: ... We were messing around.
   [But we ain't messing] [[around]] no more,
A: [Hey].
B: [Mhm]?
S:                          [[All right]].
B:                          [[Hm]].
```

Note that the use of single versus double brackets here to distinguish between two separate cases of overlap in close succession provides the crucial robustness that mere vertical alignment cannot. Even if the words "All right" (in line 6) were accidentally shifted to the left margin, the fact that they overlap with "around" and not with "Hey" could be reliably recovered.

3.4 Economy

MAKE REPRESENTATIONS ECONOMICAL. The most important reasons to build economy into the design of a discourse transcription system are efficiency in transcribing and ease of reading, with decreased data storage requirements being a lesser side benefit.

14. *Avoid verbose notations.* A short notation is generally preferred over a long one. This minimizes the number of keystrokes the transcriber must use, decreases storage demands, and improves readability by keeping the words from getting lost in the clutter.

Of course, this principle must be balanced against the need for mnemonic aids to learning (principle 7). For example, in DT whispered speech is written with two letters (**<WH WH>**) where an arbitrary one-character notation would do as well in theory;[10] but **<WH WH>** is easier to remember as whispering than, say, an arbitrary **<8 8>**.

(21) ((AFRICA))

```
A: Leopards will be hard to get.
   Leopards and cheetah.
   They're <WH fast WH>.
   ... and they hide.
   You don't see them very often.
```

On the other hand, **<WHISPER WHISPER>**, which would obviously be still easier to learn and remember, is rejected as too verbose: the numerous characters take longer to write and, by cluttering the transcription unnecessarily, make it harder to focus on the words spoken.

15. *Use short notations for high frequency phenomena.* Since the non-alphabetic symbols that have the most obvious potential for brief, non-verbose notations are a limited good, the designer should exploit them wisely. Overall economy of characters in a body of transcriptions is promoted if the briefest notations are reserved for representing things that happen most often in discourse; infrequently occurring phenomena can be written using longer notations. This can be thought of as turning Zipf's law on its head—making it a prescriptive injunction, in order to ensure that transcription data resembles natural language data in frequency distribution. For example, because laughter is generally a high frequency phenomenon in conversation, the use of the one-character @ sign for laughter in DT instead of the word (LAUGH) is not only economical in itself, it also makes for a greater total economy than assigning the @ to something that happens more rarely, such as a cough (which DT is content to write out as (COUGH)).[11] Similarly, it happens rather often that some words on a recording cannot be heard clearly; rather than write out something like "2 SYLLABLES INDECIPHERABLE", or even just ((2 sylls)),[12] an economy is gained in the DT practice of simply writing one capital letter X for each indecipherable syllable.[13] The advantage is not merely that fewer keystrokes are needed during the transcribing process,[14] but also that the reader is not bogged down by large amounts of clutter.

In addition, one of the most important functions of short notations is their potential for diagrammatic iconicity, since they can easily be iterated when the events they represent are iterated (e.g. laughter syllables or indecipherable syllables; see principle 6).

The short notations principle applies even for smaller scale differences in formal markedness. For example, audible inhalation (written as (H) in DT) is

substantially more frequent in discourse than audible exhalation (Hx); hence it is the exhalation that should be assigned the longer, "marked" notation.[15]

(22) ((AFRICA))
D: Do you think she fell out of fear?
A: **(H)** No=,

(23) ((DEPRESSION))
B: ...(4.3) **(Hx)** ... Kids in the city miss so much.

16. *Use discriminable notations for word-internal phenomena.* The reason for this is obvious: if users (and their computational tools) are to be able to recognize the lexical words in a transcription, they need to be able to filter out any symbols that appear in the middle of them. Lengthening of sounds, laughing while speaking, and overlap beginning and ending are some of the phenomena that may need to be written wherever they happen, even if this is in the middle of a word. As long as the notations for such phenomena use only non-alphabetic symbols, or employ brackets to enclose any alphabetic characters inserted within the word, they will allow unambiguous differentiation between lexical word and embedded features. For example, DT uses the @ sign for laughter (cf. example 19) and the equal sign for prosodic lengthening (example 24). Both phenomena frequently occur in the middle of words, but their notations allow them to be easily discriminated from the word itself.

(24) ((RANCH))
R: the horse is always we=t,
and it's always moi=st,
it's always on something moi=st,
... Sure it's going to be softer.

17. *Minimize word-internal notations.* Because readers tend to recognize whole words at one go by taking in their overall gestalt shape, it is important to disrupt this familiar shape as little as possible. For phenomena that must be written word-internally (such as prosodic lengthening of individual sounds), the notation should be not merely discriminable, but short. Longer notations will separate the beginning from the ending portion of a word, and make it more difficult to perceive its gestalt unity at a glance. At normal reading speed it is more difficult to assimilate something written as "both(LENGTH)er" (or even "both= = = = =er") than as "both=er". Fortunately, the non-alphabetic characters that recommend themselves on the ground of discriminability also make good short notations, requiring just a single character.

The minimum disruption of a word's sequence of letters is actually zero characters: if a notation can be moved from inside a word to its margins, this lessens interference with gestalt word recognition. For example, in DT tone is generally written before the whole *word* that bears the accent ("/\reduce"), rather than before the particular vowel on which this prominence is realized ("red/\uce") (as in the computerized version of the London-Lund Corpus; Svartvik and Quirk 1980:17). This is possible because in English as in many languages the specific location of the prominence within the word is generally predictable on lexical grounds, i.e. from information available in the dictionary.

Although it could be argued that simply enclosing alphabetic word-internal notations within parentheses achieves the end of *discriminability* (see principle 16 above) about as well as using nonalphabetic characters, the practice fares poorly on the criterion of minimizing disruption of gestalt word recognition, since it introduces at least two additional characters (the parentheses) beyond the number required for the distinctive notation itself. Thus a laugh in the middle of a word written with alphabetic symbols as "heh", if provided with the parentheses needed to unambiguously differentiate it from the surrounding letters of the word it is embedded in, will interpose five characters between the start and the finish of the word, as against one for @.

18. *Use space meaningfully.* Meaning is carried not just by characters but also by their spatial arrangement within the frame provided by line and page (Ochs 1979:45ff). If exploited effectively as a design feature, this can both economize characters and promote iconicity. The most pervasive use of space in discourse transcription is for space-time iconicity in the portrayal of temporal sequencing and simultaneity. Based on conventions of reading order traditional in most Western languages, notations written farther to the left within the line generally represent earlier events than those written farther to the right, and lines which appear higher on the page represent earlier events than lines which appear lower. (See Ehlich (forthcoming) for a different but effective approach to space-time iconicity based on musical notation.) But the complexity of multi-party interaction sometimes overwhelms this simple iconicity, as when overlapping speech by two speakers must be written on two separate lines: for practical reasons one line has to be written above the other despite there being no basis for this in the temporal meaning carried by the vertical dimension (cf. examples 1, 2, 14, 20). Even here, the horizontal dimension can still be used iconically to represent simultaneity of the overlap, as in DT and other systems. For instance, in example 16 above the brackets are placed within the sequence of several @ signs (representing extended laughter) in order to display iconically just where within the laughter sequence the overlaps begin and end.

Note that space must be used with caution, since it *can* be a problematic symbolic resource due to potential fragility (cf. principle 13). The safest, most robust spatial notations revolve around the line of text (i.e. either horizontal position within a line or vertical position in a sequence of lines), because lines are anchored by a stable, standard character—the carriage return—which, though normally invisible in itself, is readily visible in its consequences.

Space can also be exploited to indicate unit structure, as when line breaks represent intonation unit boundaries (see principle 22 below).

Another key use of space, often overlooked, is for the assignment of significance to positions within the line or page as the stable locus for a particular information category (cf. Ochs 1979). For example, in DT the far left margin consistently carries information on speaker identity;[16] the next column over is where intonation units begin (except overlapping ones); and the end of the line is always the place to look for certain intonational information. Attaching this kind of significance to a syntagmatic position, while perhaps not robust enough by itself to serve as the sole marker of a contrast, can be a big help to the reader who wishes to scan a transcription for, say, recurrent patterns in terminal pitch movement, or intonation unit truncation, or accentuation of intonation unit or turn beginnings, and so on. The common practice of simply running words on until they fill up the line wastes a valuable symbolic resource, since neither the right margin nor the line itself carries any consistent meaning. Transcription users should not have to miss out on the kind of intuitive visual organization of transcription information that a stable information locus can provide.

3.5 Adaptability

MAKE THE SYSTEM ADAPTABLE. Because spoken discourse is complex enough in its many layers that it virtually demands to be approached from a variety of viewpoints, general discourse transcription systems must be adaptable.

19. *Allow for seamless transition between degrees of delicacy.* For practical reasons a discourse researcher may begin by making a fairly broad transcription, and later decide to add additional detail. Also, it is frequently desirable to be able to present a given speech event in greater or lesser detail depending on the occasion, which may reflect the level of the audience, the kind of phenomenon being examined, and so on. For these reasons the transcription system should facilitate moving from one level of delicacy to the next, in either direction, within the same transcriptional framework and even among versions of the same text. The most important design issue here is to insure in advance that if "narrow"

notations are later added alongside the notations of an initially broad transcription, they neither introduce ambiguities nor require changes to the original transcription, which could easily introduce errors. Notational paradigms must be set up in advance to include distinctive symbols for narrow features, even if these are not as yet being used. (Less crucially, it is also useful if broad and narrow features are sufficiently well distinguished that a narrow transcription can be automatically converted to a broad one by stripping out the narrow notations.)

For example, a transcriber might begin by indicating only medium and long pauses, considering short pauses (200 milliseconds or less) as being too detailed for current purposes. In the following example, a broad transcription (using DT) shows one medium pause in the fourth line, represented by three dots:

(25) ((RANCH))

```
R: a reining pattern is,
   a pattern where you do sliding stops,
   spins,
   ... lead changes,
   I know you probably don't know what that is.
```

But if the researcher later decides to focus more attention on speech rhythms, and hence increases delicacy to the point where even short pauses are shown, all that will be required is to listen again and time the relevant (short) pauses, and then mark them where they occur. Given proper transcription system design, the increased delicacy will be accommodated smoothly: newly introduced short-pause notations (e.g. two-dot clusters in DT) will not be confused with, nor require changes in, the original representations of longer pauses.

(26) ((RANCH))

```
R: .. a reining pattern is,
   .. a pattern where you .. do sliding stops,
   .. spins,
   ... lead changes,
   .. I know you probably don't know what that is.
```

20. *Allow for seamless integration of user-defined transcription categories.* It is impossible for the designer of a general-use transcription system to anticipate everything that its users will want to do with it. The system should be designed in advance to let users introduce their own transcriptional categories, without compromising the integrity of data written using the original transcription conventions. The simplest way to do this is to leave some symbols undefined so that users can assign their own meanings to them, and also to provide for the

construction of open-ended sets of complex symbols. For example, in DT the characters " ~ ; are reserved for user-defined categories. Also, single parentheses can be used with any brief description to indicate new categories of nonverbal vocal sounds (e.g. **(SNORT)**, **(SNIFFLE)**), while angle brackets can be combined with a brief annotation to produce as many new categories as are needed for representing phenomena which apply over a stretch of speech (e.g. **<WH WH>**).

21. *Allow for seamless integration of presentation features.* Advance planning is required to ensure that notational resources will be available to people who want to highlight particular speech event features—for example, to call attention to the backchannels in a passage cited in an article on this topic. This is a fairly simple matter, since highlighting is primarily a matter of display within a limited, controlled context such as a printed article. Because the display format chosen for a publication will not ordinarily affect the integrity of the database from which its illustrative examples are drawn, resources which are too problematic for discourse transcriptions and databases per se (principles 11 and 13)—boldface, italics, underlining, and so on—can be freely used for presentation (as exemplified in the use of boldface throughout this article).

22. *Allow for seamless integration of indexing information.* Working with discourse often requires making reference to a particular place in the text, whether to attach analytical coding to a specific unit or to informally describe a discourse feature of interest. One common way to add indexing to a transcription is to attach a unique number (or complex of numbers) to each unit in the transcription, such as a line.

(27) ((CARSALES))

```
276  D: we have the used Dodge,
277     but no new ones.
278  G: ... No new ones?
```

While the unit in question can be arbitrarily delineated (for example, as the set of words that happen to fit on one line), there are important advantages if the unit is meaningful, and its boundaries significant (see principle 18 above). This allows the unit of indexing to be a significant unit of coding, which can make management of integrated transcription and coding data much easier. This is one reason a number of discourse researchers find it useful to place each intonation unit on a separate line, allowing each to be individually indexed and subsequently coded and manipulated consistently in all database operations (Du Bois and Schuetze-Coburn, forthcoming). The one-unit-per-line convention is so valuable that even when one must deal with long intonation units (or long representations

of them, as is often the case when the original text line is accompanied by a morpheme-by-morpheme gloss), it is worth taking extra measures such as using a smaller font or printing along the long (landscape) axis of the paper in order to get each unit onto a single line.

23. *Allow for seamless integration of user-defined coding information.* Because the kinds of analytical coding that discourse researchers do vary widely, transcription systems must allow for diverse user-defined coding schemes. There are two main ways to integrate coding information with transcription information. The first is to embed codes directly into the transcription; for this purpose some symbols should be reserved for user-defined coding categories (parallel to reserving symbols for user-defined transcription categories, principle 20). The second is simply to provide an indexing scheme that can access various kinds of units that may be of interest to the researcher (principle 22). With a good indexing scheme in place, the researcher can attach an unlimited amount of coding to each unit within the comfortable framework of a discourse database, without cluttering up the transcription, and without undermining its integrity with coding- rather than transcription-level analytical decisions. (For a description of a grammatical- and prosodic-unit indexing and coding scheme specifically designed for discourse research databases, see Du Bois and Schuetze-Coburn, forthcoming.)

4 Conclusions

What I have sought to do in this article is to outline some of the principles which in my view should govern the design of an effective discourse transcription system. I began by considering how discourse transcriptions are used, in order to determine what the goals of transcription design should be; and then went on to offer suggestions on the principles and conventions that ensure that transcriptions will do what is expected of them. Throughout I have been guided by an underlying philosophy of the ideal of discourse transcription: to present the multifaceted flux of discourse in a way that is as accessible to the analyst as it is to the participant. Though this ideal may represent more than any discourse transcription can ultimately hope to achieve, still a transcription can and should go a long way toward making the speech event accessible to the discourse researcher. The raw reality of discourse must be interpreted and analyzed into categories which reveal its living structure, and which are displayed in a way that is both evident to the reader's eye and accessible to any computational tools that may be called into play. The transcription design process thus takes into consideration both the nature of the speech event and the nature of discourse research, forging a structured language which articulates the apprehension of the former by the latter. This process requires effective design principles, if

transcriptions are to live up to the multiplex demands they face. Making design principles explicit provides a framework for evaluating transcription systems. It contributes to the design of new systems and the improvement of old ones, as well as to a general understanding of the role transcriptions play in shaping analyses of discourse and the theories to which they give rise.

Along with presenting some general design principles, I have made frequent reference to one particular transcription system. The *Discourse Transcription* system has grown out of the work of many individuals representing a variety of approaches to the study of discourse, as codified and further developed by myself and my colleagues (Du Bois et al. forthcoming a, b). This system, or variants of it, is now being used by a growing number of researchers at various universities around the world to pursue a broad range of questions about spoken discourse—questions about how speakers use features like discourse particles, sentence structure, and intonation to organize the flow of information and interaction in conversation; how speakers achieve their interactional goals through speech; how social groups differ in their conversational styles; how spoken genres, such as conversations, political speeches, sermons, and so on, differ from each other; how spoken language differs from written language; how languages with sharply distinct typological features manage the same discourse functions; and many other questions. Because this system is designed to be used for such diverse purposes, we have worked hard to make sure it adequately represents a broad range of discourse features, within an overall framework that is coherent yet adaptable.

In fact, the DT transcription system is best seen simply as a framework within which representations of speech events can be created—a framework designed from the outset to allow for creative adaptation and growth. There are several dimensions along which further development can be hoped for in coming years—for example, toward a precise but readable interweaving within the transcription of representations of speech rhythms and timing; detailed pitch and amplitude information; and nonverbal cues like eye gaze, body orientation, and so on. Inevitably, as new research directions arise in coming years there will be new and unforeseen transcriptional functions to be incorporated, and old conventions to be adapted. To the extent that this or any other system of discourse transcription must develop and adapt during the whole of its active lifetime, it is to be hoped that the design principles articulated here will contribute to the next round of advances.

NOTES

1. This paper is based on a series of presentations made during September 1989 at Uppsala and Lund Universities and at the Stockholm Conference on the Use of Computers in Language Teaching and Research; and, at UC Santa Barbara, at the Conference on Current Issues in Corpus Linguistics, June 1990. I am especially indebted to Stephan Schuetze-Coburn, Jane Edwards and Ken Whistler for helpful discussions of several of the issues treated in this paper, and to Robert Potter for information on the history of playwrights' notational conventions. I would also like to express my appreciation for helpful comments on earlier incarnations of this work to Wallace Chafe, Alan Cruttenden, Susanna Cumming, Alessandro Duranti, Geoffrey Leech, Marianne Mithun, Elinor Ochs, Danae Paolino, Stig Johanssen, Emanuel Schegloff, Jan Svartvik, Gunnel Tottie, Sandra Thompson, and many others who have made invaluable suggestions during the long evolution of our transcription system. They are of course not accountable for my decisions in the end.

2. Nor is this by any means to endorse the doctrine once known as separation of levels — now seeking rehabilitation under the label of modularity, i.e. internal mutual autonomy (non-integration) of compartmentalized system components.

3. Except as otherwise noted, examples throughout this article will be presented in a fairly broad transcription (Du Bois et al. forthcoming a). They will introduce more detailed "narrow" features only as necessary for illustrating specific transcriptional categories.

4. All-capitals are also used for other kinds of non-speech, such as speaker identification labels.

5. This iconic notation, incorporated into DT, also gains support from modern scholarly tradition, given the use of similar notations (e.g. arrows or lines sloping upward-right versus downward-right) by many others including Crystal (1969), Svartvik and Quirk (1980:17, 22), and the International Phonetic Association (1989).

6. The fact that the iconicity doesn't gain strength until the conventional meaning is revealed may convince some that it is not "real" iconicity. But as Whorf, Ullmann and others long ago observed, such after-the-fact iconicity may indeed be the norm, and in any case has important consequences for how speakers understand and use the symbol in question.

7. There is hope in the not-too-distant future for an international standard for computers (including microcomputers) that would encompass more than 64,000 characters, including virtually all of the world's writing systems (Ken Whistler,

personal communication). If this standard comes into universal use, the constraints on the selection of symbols for discourse transcription systems will ease considerably. However, the fundamental design principles outlined here will remain valid.

8. But see principle 21 for legitimate uses of such notational resources.

9. Another contrast which, though less problematic, is not entirely reliable, is that between lower-case and upper-case letters, when the notation is confronted with words which are obligatorily written entirely in capitals, such as the English first person 'I' or commonly pronounced acronyms like 'USA', 'ACLU', and so on.

10. Since DT uses **<W W>** as a mnemonic (and non-verbose) notation for "widened" pitch movement, the letter W is not available for use here.

11. Of course, sounds don't come already analyzed into categories: the transcriber must actively classify the continuous diversity of human vocal expression if such discrete transcriptional categories as "laugh" and "cough" are to be employed. This is true whether the categories are represented in single characters like "@" or in strings like "COUGH", whose apparent orthographic success in capturing the reality of a "cough" is of course only illusory. It remains true whether the categories are few and broad (laugh, cough, etc.) or many and narrow (distinguishing several types of laughter, for example). Spelling out category labels is a mnemonic aid, not a depiction of natural fact.

12. See Svartvik and Quirk (1980:24).

13. The letter **X** has in the past been used in child language transcription to indicate exact repetition of an utterance (Ochs 1979:65), but this practice seems less common nowadays.

14. Assuming that the indecipherable stretches will normally be relatively short, of course.

15. But see Jefferson's system (Atkinson and Heritage 1984), which reverses this markedness relation with **.h** and **h**, respectively.

16. The meaningful use of space is a key characteristic of play- and film-scripts, though specific conventions vary (cf. examples 4 and 11 above, and Catron (1984)).

REFERENCES

Atkinson, J. Maxwell and John Heritage (eds.) (1984) Transcript notation. *Structures of social action: Studies in conversation analysis*. Cambridge: Cambridge University Press, ix-xvi.

Catron, Louis E. (1984) *Writing, producing, and selling your play*. Englewood Cliffs, New Jersey: Prentice-Hall.

Chafe, Wallace L. (1987) Cognitive constraints on information flow. In Russell S. Tomlin (ed.), *Coherence and grounding in discourse*. Amsterdam: Benjamins, 21-52.

Crystal, David (1969) *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Disraeli, Benjamin (1904) *Sybil; or, The two nations*. London: M. Walter Dunne.

Du Bois, John W. (1985) Competing motivations. In John Haiman (ed.), *Iconicity in syntax*. Typological studies in language, vol. 6. Amsterdam: Benjamins, 343-65.

Du Bois, John W. (1986) The pragmatics of marginality: Illocutionless language in cartoon semiotics. Paper delivered to the Conference on Sound Symbolism, UC Berkeley.

Du Bois, John W. (forthcoming) Typology of discourse information. MS, UC Santa Barbara.

Du Bois, John W. and Stephan Schuetze-Coburn (forthcoming) Representing hierarchy: Constituent structure for discourse databases. In Edwards and Lampert (forthcoming).

Du Bois, John W., Stephan Schuetze-Coburn, Danae Paolino, and Susanna Cumming (forthcoming a) *Discourse transcription*. MS, UC Santa Barbara.

Du Bois, John W., Stephan Schuetze-Coburn, Danae Paolino, and Susanna Cumming (forthcoming b) Outline of discourse transcription. In Edwards and Lampert (forthcoming).

Edwards, Jane A. (1989) *Transcription and the new functionalism: A counterproposal to CHILDES' CHAT conventions.* Berkeley Cognitive Science Report No. 58, March 1989. Institute of Cognitive Studies, University of California, Berkeley.

Edwards, Jane A. (forthcoming) Transcription in discourse. In William Bright (ed.), *Oxford International Encyclopedia of Linguistics.* Oxford: Oxford University Press.

Edwards, Jane A. and Martin D. Lampert (eds.) (forthcoming) *Talking data: Transcription and coding of spoken discourse.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Ehlich, Konrad (forthcoming) HIAT—A transcription system for discourse data. In Edwards and Lampert (forthcoming).

Hillerman, Tony (1990) *Coyote waits.* Harper and Row.

International Phonetic Association (1989) Report on the 1989 Kiel Convention. *Journal of the International Phonetic Association* 19: 67-80.

Ladefoged, Peter (1990) Some reflections on the new IPA. *UCLA Working Papers in Linguistics* 74: 61-76.

Lodge, David (1984) *Small World.* New York: Warner.

Ochs, Elinor (1979) Transcription as theory. In Elinor Ochs and Bambi B. Schieffelin (eds.), *Developmental pragmatics.* New York: Academic Press, 43-72.

Shakespeare, William (1623 [1928]) *The tempest. A facsimile of the first folio text.* London: Faber and Faber.

Wilde, Oscar (1982) The importance of being earnest. *The annotated Oscar Wilde.* New York: Clarkson Potter.

Williams, Tennessee (1971) The glass menagerie. *The theatre of Tennessee Williams, vol. 1.* New York: New Directions.

Svartvik, Jan and Randolph Quirk (eds.) (1980) *A corpus of English conversation*. Lund Studies in English 56. Lund: C.W.K. Gleerup.

APPENDIX 1. Symbols for Discourse Transcription

UNITS	
Intonation unit	*{carriage return}*
Truncated intonation unit	--
Word	*{space}*
Truncated word	-
SPEAKERS	
Speaker identity/turn start	:
Speech overlap	[ ]
TRANSITIONAL CONTINUITY	
Final	.
Continuing	,
Appeal	?
TERMINAL PITCH DIRECTION	
Fall	\
Rise	/
Level	–
ACCENT AND LENGTHENING	
Primary accent	^
Secondary accent	‘
Booster	!
Lengthening	=
TONE	
Fall	\
Rise	/
Fall-rise	\/
Rise-fall	/\
Level	–
PAUSE	
Long	...(N)
Medium	...
Short	..
Latching	(0)
VOCAL NOISES	
Vocal noises	( )
Inhalation	(H)
Exhalation	(Hx)
Glottal stop	%
Laughter	@

QUALITY	
Quality	<Y Y>
Laugh quality	<@ @>
Quotation quality	<Q Q>
Multiple quality features	<Y<Z Z>Y>
PHONETICS	
Phonetic/phonemic transcription	(/ /)
TRANSCRIBER'S PERSPECTIVE	
Researcher's comment	(( ))
Uncertain hearing	<X X>
Indecipherable syllable	X
SPECIALIZED NOTATIONS	
Duration	(N)
Intonation unit continued	&
Intonation subunit boundary	|
Embedded intonation unit	<| |>
Reset	*{Capital Initial}*
False start	< >
Codeswitching	<L2 L2>
NON-TRANSCRIPTION LINES	
Non-transcription line	$
Interlinear gloss line	$G
RESERVED SYMBOLS	
Phonemic/orthographic	'
Morphosyntactic coding	+ * # { }
User-definable	" ~ ;

APPENDIX 2. Synopsis of Transcription Design Principles

CATEGORY DEFINITION: *Define good categories.*

1. Define transcriptional categories which make the necessary distinctions among discourse phenomena.
2. Define sufficiently explicit categories.
3. Define sufficiently general categories.
4. Contrast data types.

ACCESSIBILITY: *Make the system accessible.*

5. Use familiar notations.
6. Use motivated notations.
7. Use easily learned notations.
8. Segregate unfamiliar notations.
9. Use notations which maximize data access.
10. Maintain consistent appearance across modes of access.

ROBUSTNESS: *Make representations robust.*

11. Use widely available characters.
12. Avoid invisible contrasts.
13. Avoid fragile contrasts.

ECONOMY: *Make representations economical.*

14. Avoid verbose notations.
15. Use short notations for high frequency phenomena.
16. Use discriminable notations for word-internal phenomena.
17. Minimize word-internal notations.
18. Use space meaningfully.

ADAPTABILITY: *Make the system adaptable.*

19. Allow for seamless transition between degrees of delicacy.
20. Allow for seamless integration of user-defined transcription categories.
21. Allow for seamless integration of presentation features.
22. Allow for seamless integration of indexing information.
23. Allow for seamless integration of user-defined coding information.